



Original Investigation | Substance Use and Addiction

Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions

Wei-Hsuan Lo-Ciganic, PhD; James L. Huang, PhD; Hao H. Zhang, PhD; Jeremy C. Weiss, MD, PhD; Yonghui Wu, PhD; C. Kent Kwok, MD; Julie M. Donohue, PhD; Gerald Cochran, PhD; Adam J. Gordon, MD, MPH; Daniel C. Malone, PhD; Courtney C. Kuza, PhD; Walid F. Gellad, MD, MPH

Abstract

IMPORTANCE Current approaches to identifying individuals at high risk for opioid overdose target many patients who are not truly at high risk.

OBJECTIVE To develop and validate a machine-learning algorithm to predict opioid overdose risk among Medicare beneficiaries with at least 1 opioid prescription.

DESIGN, SETTING, AND PARTICIPANTS A prognostic study was conducted between September 1, 2017, and December 31, 2018. Participants (n = 560 057) included fee-for-service Medicare beneficiaries without cancer who filled 1 or more opioid prescriptions from January 1, 2011, to December 31, 2015. Beneficiaries were randomly and equally divided into training, testing, and validation samples.

EXPOSURES Potential predictors (n = 268), including sociodemographics, health status, patterns of opioid use, and practitioner-level and regional-level factors, were measured in 3-month windows, starting 3 months before initiating opioids until loss of follow-up or the end of observation.

MAIN OUTCOMES AND MEASURES Opioid overdose episodes from inpatient and emergency department claims were identified. Multivariate logistic regression (MLR), least absolute shrinkage and selection operator-type regression (LASSO), random forest (RF), gradient boosting machine (GBM), and deep neural network (DNN) were applied to predict overdose risk in the subsequent 3 months after initiation of treatment with prescription opioids. Prediction performance was assessed using the C statistic and other metrics (eg, sensitivity, specificity, and number needed to evaluate [NNE] to identify one overdose). The Youden index was used to identify the optimized threshold of predicted score that balanced sensitivity and specificity.

RESULTS Beneficiaries in the training (n = 186 686), testing (n = 186 685), and validation (n = 186 686) samples had similar characteristics (mean [SD] age of 68.0 [14.5] years, and approximately 63% were female, 82% were white, 35% had disabilities, 41% were dual eligible, and 0.60% had at least 1 overdose episode). In the validation sample, the DNN (C statistic = 0.91; 95% CI, 0.88-0.93) and GBM (C statistic = 0.90; 95% CI, 0.87-0.94) algorithms outperformed the LASSO (C statistic = 0.84; 95% CI, 0.80-0.89), RF (C statistic = 0.80; 95% CI, 0.75-0.84), and MLR (C statistic = 0.75; 95% CI, 0.69-0.80) methods for predicting opioid overdose. At the optimized sensitivity and specificity, DNN had a sensitivity of 92.3%, specificity of 75.7%, NNE of 542, positive predictive value of 0.18%, and negative predictive value of 99.9%. The DNN classified patients into low-risk (76.2% [142 180] of the cohort), medium-risk (18.6% [34 579] of the cohort), and high-risk (5.2% [9747] of the cohort) subgroups, with only 1 in 10 000 in the low-risk subgroup having an

(continued)

Key Points

Question Can machine-learning approaches predict opioid overdose risk among fee-for-service Medicare beneficiaries?

Findings In this prognostic study of the administrative claims data of 560 057 Medicare beneficiaries, the deep neural network and gradient boosting machine models outperformed other methods for identifying risk, although positive predictive values were low given the low prevalence of overdose episodes.

Meaning Machine-learning algorithms using administrative data appear to be a valuable and feasible tool for more accurate identification of opioid overdose risk.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Open Access. This is an open access article distributed under the terms of the CC-BY License.

Abstract (continued)

overdose episode. More than 90% of overdose episodes occurred in the high-risk and medium-risk subgroups, although positive predictive values were low, given the rare overdose outcome.

CONCLUSIONS AND RELEVANCE Machine-learning algorithms appear to perform well for risk prediction and stratification of opioid overdose, especially in identifying low-risk subgroups that have minimal risk of overdose.

JAMA Network Open. 2019;2(3):e190968.

Corrected on July 5, 2019. doi:[10.1001/jamanetworkopen.2019.0968](https://doi.org/10.1001/jamanetworkopen.2019.0968)

Introduction

In 2016, 11.8 million American individuals reported using prescription opioids nonmedically,¹ and an estimated 115 individuals died each day from opioid overdose.²⁻⁴ The annual cost of misuse or abuse of opioids exceeds \$78.5 billion, including the costs of health care, lost productivity, substance abuse treatment, and the criminal justice system.⁵

In response, health systems, payers, and policymakers have developed programs to identify and intervene in individuals at high risk of problematic opioid use and overdose. These programs, whether outreach calls from case managers, prior authorizations, referrals to substance use disorder specialists, dispensing of naloxone hydrochloride, or enrollment in lock-in programs, can be expensive to payers and burdensome to patients. The determination of who is at high risk is a factor in the size and scope of these interventions and the resources expended. Yet, the definition of high risk is variable, ranging from a high-dose opioid (defined using various cut points) to the number of pharmacies or prescribers that a patient visits. These criteria, for example, determine how Medicare beneficiaries are selected into so-called lock-in programs in Medicare, also called the Comprehensive Addiction and Recovery Act (CARA) drug management programs.⁶ These programs will soon be required for all Part D plans.⁷

These current measures of high risk were derived from studies that used traditional statistical methods to identify risk factors for overdose rather than predict an individual's risk.⁸⁻³¹ However, individual risk factors may not be strong predictors of overdose risk.³² Moreover, traditional statistical approaches have limited ability to handle nonlinear risk prediction and complex interactions among predictors. For example, receipt of a high-dose opioid is a well-known overdose risk factor, but the complex interactions between opioid dose, substance use disorders, mental health, emergency department visits, prescriber characteristics, and socioeconomic variables may yield greater predictive power than one factor alone. The few previous studies focused on predicting opioid overdose (rather than simply identifying risk factors) either had suboptimal prediction performance^{24,28,31} or used case-control designs that were unable to measure true overdose incidence and may not adequately calibrate algorithms to real data for rare outcomes such as overdose.^{22,25,30}

Machine learning is an alternative analytic approach to handling complex interactions in large data, discovering hidden patterns, and generating actionable predictions in clinical settings. In many cases, machine learning is superior to traditional statistical techniques.³³⁻³⁸ Machine learning has been widely used in activities from fraud detection to genomic studies but, to our knowledge, has not yet been applied to address the opioid epidemic. Our overall hypothesis was that a machine-learning algorithm would perform better in predicting opioid overdose risk compared with traditional statistical approaches.

The objective of this study was to develop and validate a machine-learning algorithm to predict opioid overdose among Medicare beneficiaries with at least 1 opioid prescription. Based on the prediction score, we stratified beneficiaries into subgroups at similar overdose risk to support clinical decisions and improved targeting of intervention. We chose Medicare because of the high prevalence

of prescription opioid use and the availability of national claims data and because the program will require specific interventions targeting individuals at high risk for opioid-associated morbidity.^{6,7}

Methods

Design and Sample

This prognostic study was conducted between September 1, 2017, and December 31, 2018. The University of Arizona Institutional Review Board approved the study. This study followed the Standards for Reporting of Diagnostic Accuracy (STARD) and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines.^{39,40}

We included prescription drug and medical claims for a 5% random sample of Medicare beneficiaries between January 1, 2011, and December 31, 2015. We identified fee-for-service adult beneficiaries without cancer who were US residents and received 1 or more opioid prescriptions during the study period. We excluded beneficiaries who (1) filled only parenteral opioid prescriptions and/or cough or cold medication prescriptions containing opioids, (2) had malignant cancer diagnoses (eTable 1 in the [Supplement](#)), (3) received hospice, (4) ever enrolled in Medicare Advantage plans (because their health care use in Medicare Advantage may not be observable), or (5) had their first opioid prescription after October 1, 2015 (eFigure 1 in the [Supplement](#)). An index date was defined as the date of a patient's first opioid prescription between April 1, 2011, and September 30, 2015. Once eligible, beneficiaries remained in the cohort, regardless of whether they continued to receive opioid prescriptions, until they were censored because of death or the end of observation.

Outcome Variables: Opioid Overdose

We identified any occurrence of fatal or nonfatal opioid overdose (prescription opioids or other opioids, including heroin), defined in each 3-month window after the index prescription using the *International Classification of Diseases, Ninth Revision*, and *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10)*, codes for overdose (eTable 2 in the [Supplement](#)) from inpatient or emergency department settings.^{14,41-44} Overdose was defined with either an opioid overdose code as the primary diagnosis (80% of identified overdose episodes) or other drug overdose or substance use disorder code as the primary diagnosis (eTable 3 in the [Supplement](#)) and opioid overdose as the nonprimary diagnosis (20% of identified overdose episodes), as defined previously.¹⁴ Sensitivity analyses using opioid overdose as the primary diagnosis and capturing any opioid overdose diagnosis code in any position yielded similar results.

Predictor Candidates

We compiled 268 predictor candidates, informed by the literature (eTable 4 in the [Supplement](#)).⁸⁻³¹ Patient, practitioner, and regional factors were measured at baseline in the 3 months before the first opioid prescription fill and in 3-month windows after initiating prescription opioids. We chose a 3-month window in accordance with the literature and to be consistent with the quarterly evaluation period commonly used by prescription drug monitoring programs and health plans.^{13,14,45} In the primary analysis, we used the variables measured in each 3-month period (eg, the first) to predict overdose risk in each subsequent 3-month period (eg, the second) (eFigure 2A in the [Supplement](#)). In sensitivity analyses, instead of using a previous 3-month period to predict overdose in the next period, we included information collected in all of the historical 3-month windows to predict opioid risk for each 3-month period for each person (eFigure 2B in the [Supplement](#)).

The predictor candidates also included a series of variables related to prescription opioid and relevant medication use: (1) total and mean daily morphine milligram equivalent (MME),¹⁷ (2) cumulative and continuous duration of opioid use (ie, no gap >32 days between fills),⁴⁵ (3) total number of opioid prescriptions overall and by active ingredient, (4) type of opioid based on the US Drug Enforcement Administration's Controlled Substance Schedule (I-IV) and duration of action, (5)

number of opioid prescribers, (6) number of pharmacies providing opioid prescriptions,^{11,17,23} (7) number of early opioid prescription refills (refilling opioid prescriptions >3 days before the previous prescription runs out),^{4,6} (8) cumulative days of early opioid prescription refills, (9) cumulative days of concurrent benzodiazepines and/or muscle relaxant use, (10) number and duration of other relevant prescriptions (eg, gabapentinoids), and (11) receipt of methadone hydrochloride or buprenorphine hydrochloride for opioid use disorder.^{19,47-50}

Patient sociodemographic characteristics included age, sex, race/ethnicity, disability as the reason for Medicare eligibility, receipt of low-income subsidy, and urbanicity of county of residence. Health status factors (eg, number of emergency department visits) were derived from the literature and are listed in eTable 4 in the [Supplement](#).^{13,16,51-54} Practitioner factors included opioid prescriber's sex, specialty, mean monthly opioid prescribing volume and MME, and mean monthly number of patients receiving opioids. Many beneficiaries had more than 1 opioid prescriber, in which case the practitioner prescribing the highest number of opioids was designated as the primary prescriber. Regional factors (eg, percentage of households below the federal poverty level) included variables obtained from publicly available resources, including the Area Health Resources Files, Area Deprivation Index data sets,⁵⁵ and County Health Rankings data.⁵⁶

Machine-Learning Approaches and Prediction Performance Evaluation

Our primary goal was risk prediction, and the secondary goal was risk stratification (ie, identifying patient subgroups at similar overdose risk). First, we randomly and equally divided the cohort into training (developing algorithms), testing (refining algorithms), and validation (evaluating algorithm's prediction performance) samples. In both the primary and sensitivity analyses (eFigure 2 in the [Supplement](#)), we developed and tested prediction algorithms for opioid overdose using 5 commonly used machine-learning approaches: multivariate logistic regression, least absolute shrinkage and selection operator-type regression (LASSO), random forest (RF), gradient boosting machine (GBM), and deep neural network (DNN). Previous studies consistently showed that these methods yield the best prediction results^{57,58}; the eAppendix in the [Supplement](#) describes the details for each approach used. Given that beneficiaries may have multiple opioid overdose episodes, we present the results from a patient-level random subset (ie, using one 3-month period with predictor candidates measured to predict risk in the subsequent 3 months for each patient) from the validation data for ease of interpretation. Episode-level performance was the same as the patient-level results.

To assess discrimination performance (ie, the extent to which patients who were predicted to be high risk exhibited higher overdose rates compared with those who were predicted to be low risk), we compared the C statistic (or area under the receiver operating curve) and precision-recall curves⁵⁹ across different methods from the validation sample using the DeLong Test.⁶⁰ Given that overdose events are rare outcomes and C statistics do not incorporate information about outcome prevalence, we reported other metrics, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio, negative likelihood ratio, number needed to evaluate (NNE) to identify 1 overdose episode, and estimated rate of alerts, to thoroughly assess the algorithms' prediction ability (eFigure 3 in the [Supplement](#)).^{61,62} To compare performance across methods, we presented and assessed these metrics at the optimized prediction threshold that balances sensitivity and specificity, as identified by the Youden index.⁶³ Furthermore, because no single threshold is suitable for every purpose, we also presented these metrics at multiple other levels of sensitivity and specificity (eg, arbitrarily choosing 90% sensitivity) to enable risk-benefit evaluations of potential interventions that use different thresholds defining high risk.

On the basis of the distribution of individuals' estimated probability of an overdose event, we classified beneficiaries in the validation sample into low risk (predicted score below the optimized threshold), medium risk (score between the optimized threshold and top fifth percentile), or high risk (the top fifth percentile of scores, chosen according to clinical utility). We evaluated calibration plots (the extent to which the predicted overdose risk agreed with the observed risk) by the 3 risk groups.

To ensure clinical utility, we reported the predictors with the strongest effect. Because no standardized methods exist to identify individual important predictors from the DNN model, we reported the top 50 important predictors from the GBM and RF models. We also compared our prediction performance over a 12-month period with any of the 2019 Centers for Medicare & Medicaid Services opioid safety measures, which are meant to identify high-risk individuals or utilization behavior in Medicare.⁶⁴ These simpler decision metrics were constructed from factors identified from previous studies using traditional approaches (eg, multivariate logistic regression). These measures included 3 metrics: (1) high-dose use, defined as higher than 120 MME for 90 or more continuous days; (2) 4 or more opioid prescribers and 4 or more pharmacies; and (3) concurrent opioid and benzodiazepine use for 30 or more cumulative days. In addition to using 3- and 12-month windows, we conducted a sensitivity analysis using a 6-month window in DNN to examine whether the prediction quality changes with different time horizons.⁶⁵

Statistical Analysis

We compared the patient characteristics by overdose status and by training, testing, and validation sample with unpaired, 2-tailed *t* test, χ^2 test and analysis of variance, or corresponding nonparametric tests, as appropriate. We assessed correlations between 2 variables using Pearson correlation coefficient (*r*). Statistical significance was defined as 2-tailed *P* < .05.

All analyses were performed using SAS, version 9.4 (SAS Institute Inc); Python, version 3.6 (Python Software Foundation); and Salford Predictive Modeler software suite, version 8.2 (Salford System).

Results

Patient Characteristics

Beneficiaries in the training (*n* = 186 686), testing (*n* = 186 685), and validation (*n* = 186 686) samples had similar characteristics and outcome distributions (approximately 63% were female, 82% were white, 35% had disabilities, and 41% were dual eligible; the mean [SD] age was 68.0 [14.5] years (eTable 5 in the [Supplement](#)). Overall, 3188 beneficiaries (0.6%) had at least 1 opioid overdose episode during the study period.

Prediction Performance of Machine-Learning Algorithms

Figure 1 summarizes 4 prediction performance measures of each model. The DNN (C statistic = 0.91; 95% CI, 0.88-0.93) and GBM (C statistic = 0.90; 95% CI, 0.87-0.94) algorithms outperformed the LASSO (C statistic = 0.84; 95% CI, 0.80-0.89), RF (C statistic = 0.80; 95% CI, 0.75-0.84), and multivariate logistic regression (C statistic = 0.75; 95% CI, 0.69-0.80) methods for predicting opioid overdose (*P* < .001). In addition, DNN and GBM had similar prediction performance, and DNN had the best precision-recall performance (Figure 1B), based on an area under the curve of 0.036. Sensitivity analyses including all the historical 3-month windows yielded similar results (eFigure 4 in the [Supplement](#)).

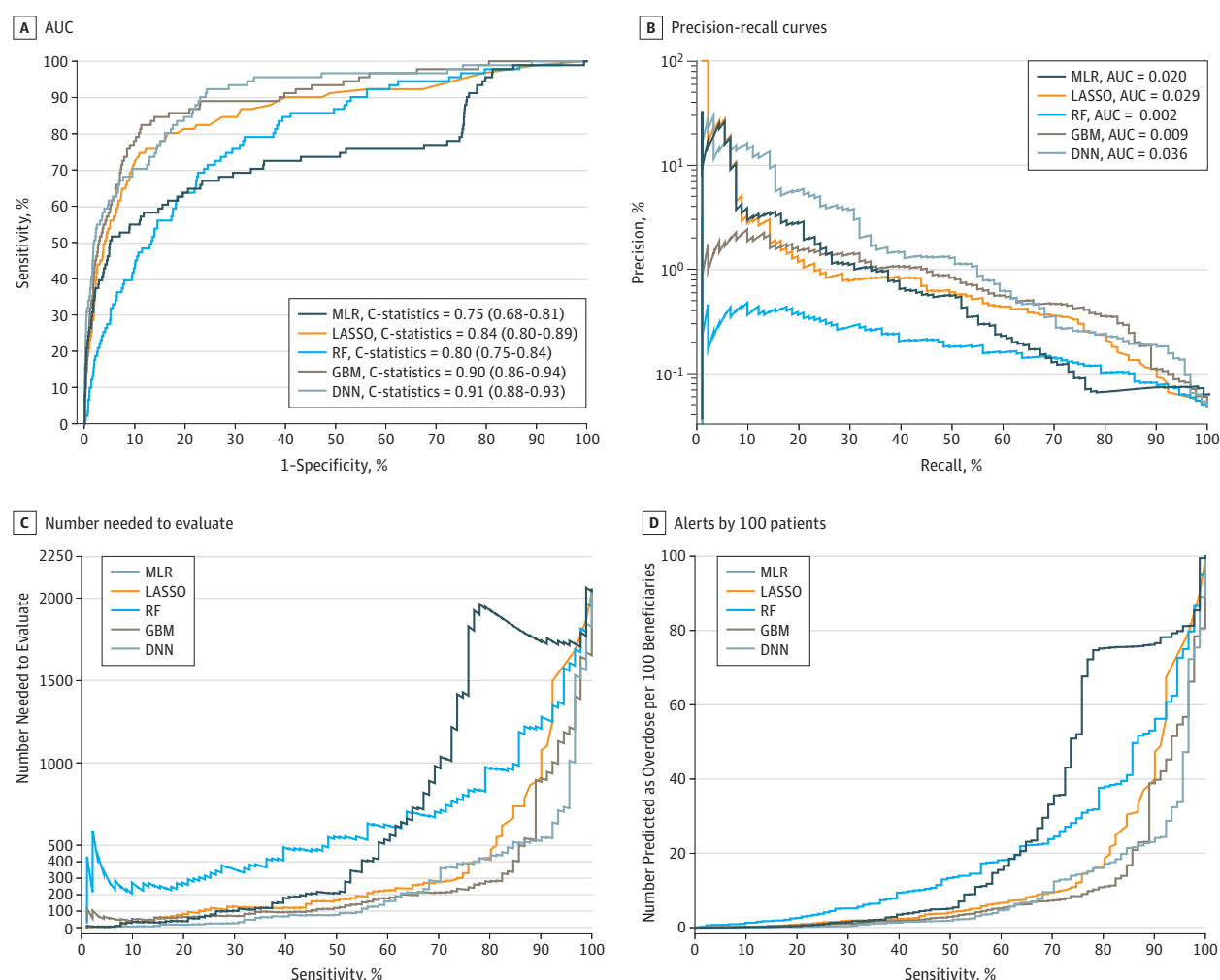
eTable 6 in the [Supplement](#) shows the prediction performance measures across different levels (90%-100%) of sensitivity and specificity for each method. At the optimized sensitivity and specificity, as measured by the Youden index, GBM had a sensitivity of 86.8%, specificity of 81.1%, PPV of 0.22%, NPV of 99.9%, NNE of 447, and 24 positive alerts per 100 beneficiaries. Similarly, at the optimized sensitivity and specificity, DNN had a sensitivity of 92.3%, specificity of 75.7%, PPV of 0.18%, NPV of 99.9%, NNE of 542, and 22 positive alerts per 100 beneficiaries (Figure 1C and D; eTable 6 in the [Supplement](#)). If sensitivity were instead set at 90% (ie, attempting to identify 90% of individuals with actual overdose episodes), GBM had a specificity of 72.3%, PPV of 0.16%, NPV of 99.9%, NNE of 631 to identify 1 overdose, and 28 positive alerts generated per 100 beneficiaries; DNN had a specificity of 77.0%, PPV of 0.19%, NPV of 99.9%, NNE of 525, and 23 positive alerts per 100 beneficiaries (eTable 6 in the [Supplement](#)). If specificity were set at 90% (ie, identifying 90%

of individuals with actual nonoverdose), GBM had a sensitivity of 74.7%, PPV of 0.41%, NPV of 99.9%, NNE of 245, and 9 positive alerts per 100 beneficiaries; DNN had a sensitivity of 70.3%, PPV of 0.34%, and NPV of 99.9%, NNE of 294, and 10 positive alerts per 100 beneficiaries. Overall, DNN's prediction scores were highly correlated with the GBM's prediction scores ($r=0.73$ for all patients, 0.73 for those without overdose episodes, 0.80 for those with overdose episodes; eFigure 5 in the Supplement).

Risk Stratification Using Predicted Probability

Using the GBM algorithm, 144 860 (77.6%) of the sample were categorized into low risk, 32 415 (17.4%) into medium risk, and 9411 (5.0%) into high risk for overdose (Table 1). Among all 91 beneficiaries with an overdose episode in the sample, 54 (59.3%) were captured in the high-risk group. Similarly, using the DNN algorithm, 9747 individuals (5.2%) were predicted to be high risk, capturing 56 overdose episodes (61.5%). Among the 142 180 individuals (76.2%) categorized as low risk, 99.99% did not have an overdose. Figure 2 depicts the actual overdose rate for individuals in

Figure 1. Performance Matrix of Machine-Learning Models for Predicting Opioid Overdose in Medicare Beneficiaries



The 4 prediction performance matrixes in the validation sample are the area under the receiver operating characteristic curve (AUC) or C statistic (A); the precision-recall curves, which have improved performance if they are closer to the upper right corner or above the other method (B); the number needed to evaluate (NNE) by different cutoffs of sensitivity (C); and alerts per 100 patients by different cutoffs of sensitivity (D).

DNN indicates deep neural network; GBM, gradient boosting machine; LASSO, least absolute shrinkage and selection operator-type regularized regression; MLR, multivariate logistic regression; and RF, random forest.

each of the 3 risk groups. Across both the GBM and DNN models, those in the high-risk group had 7 to 8 times the risk of overdose compared with those in the lower-risk groups (observed overdose rate of GBM: 0.57% [high risk], 0.08% [medium risk], and 0.01% [low risk]; observed overdose rate of DNN: 0.57% [high risk], 0.07% [medium risk], and 0.01% [low risk]). Again, depicted is the negligible rate of overdose in the low-risk subgroups, representing more than three-quarters of the sample.

Figure 3 shows the most important predictors (n = 50) identified by the GBM model, such as total MME, history of any substance use disorder, mean daily MME, age, and Medicare disability status. eFigure 6 in the Supplement shows the most important predictors (n = 50) identified by the RF model.

Table 1. Prediction Performance of Gradient Boosting Machine and Deep Neural Network Models in the Validation Sample Divided Into Risk Subgroups^a

Performance Metric	GBM			DNN		
	Low Risk	Medium Risk	High Risk	Low Risk	Medium Risk	High Risk
Total, No. (%)	144 860 (77.6)	32 415 (17.4)	9411 (5.0)	142 180 (76.2)	34 759 (18.6)	9747 (5.2)
Predicted score, median (range) ^b	14.6 (1.4-39.0)	55.4 (39.0-77.7)	83.8 (77.7-93.8)	14.2 (2.1-46.5)	61.6 (46.5-81.9)	88.7 (81.9-99.7)
No. of actual overdose episodes (% of each subgroup)	11 (0.01)	26 (0.08)	54 (0.57)	9 (0.01)	26 (0.07)	56 (0.57)
No. of actual nonoverdose episodes (% of each subgroup)	144 849 (99.99)	32 389 (99.92)	9357 (99.43)	142 171 (99.99)	34 733 (99.93)	9691 (99.43)
Sensitivity, %	0	100	100	0	100	100
PPV, % ^c	NA	0.08	0.57	NA	0.07	0.57
NNE ^c	NA	1247	174	NA	1337	174
Specificity, %	100	0	0	100	0	0
NPV, % ^c	99.99	NA	NA	99.99	NA	NA
Overall No. of misclassified overdose episodes (% of overall cohort) ^c	11 (0.006)	32 389 (17.4)	9357 (5.0)	9 (0.005)	34 733 (18.6)	9691 (5.2)
% of All overdose episodes captured over 3 mo (n = 91)	12.1	29.6	59.3	9.9	28.6	61.5

Abbreviations: DNN, deep neural network; GBM, gradient boosting machine; NA, not able to be calculated owing to 0 denominator; NNE, number needed to evaluate; NPV, negative predictive value; PPV, positive predictive value.

^a Risk subgroups were classified into low risk (score below the optimized threshold), medium risk (predicted score between the optimized threshold and the top fifth percentile score), and high risk (predicted score in the top fifth percentile). The optimized thresholds were 39 (or probability of 0.39) for GBM and 46.5 (or probability of 0.465) for DNN.

^b Predicted scores were calculated by the predicted probability of overdose multiplied by 100.

^c If classifying medium- and high-risk groups as overdose and low-risk group as nonoverdose, then the PPV and NNE were not able to be calculated for the low-risk group because this group was considered as nonoverdose. Similarly, the NPV was not able to calculate for the medium- and high-risk groups because these groups were considered as overdose. Detailed definitions of prediction performance metrics are provided in eFigure 3 in the Supplement.

Figure 2. Calibration Performance of Gradient Boosting Machine (GBM) and Deep Neural Network (DNN) by Risk Group

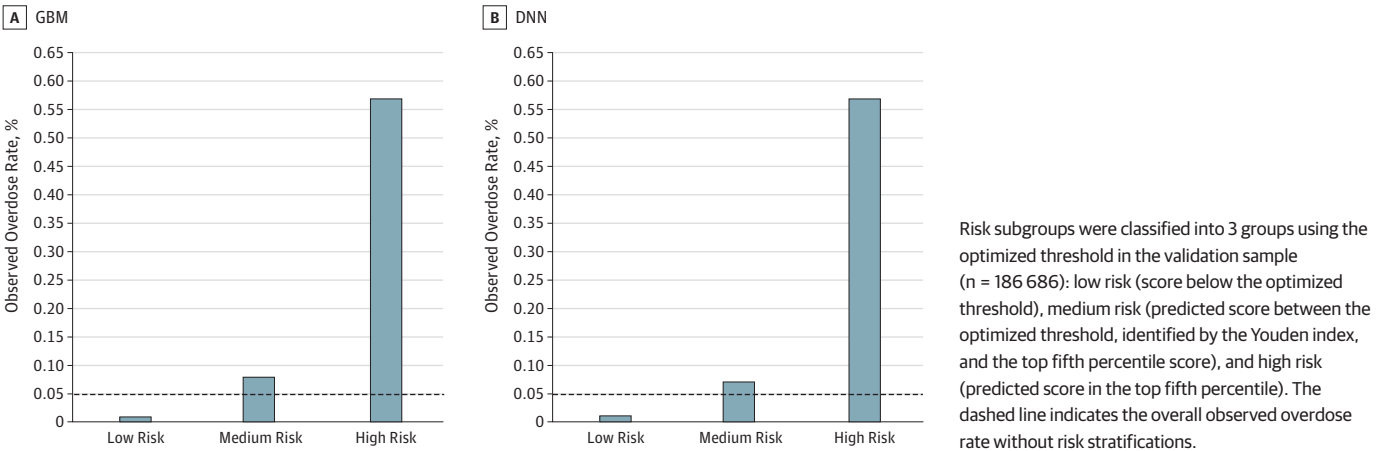
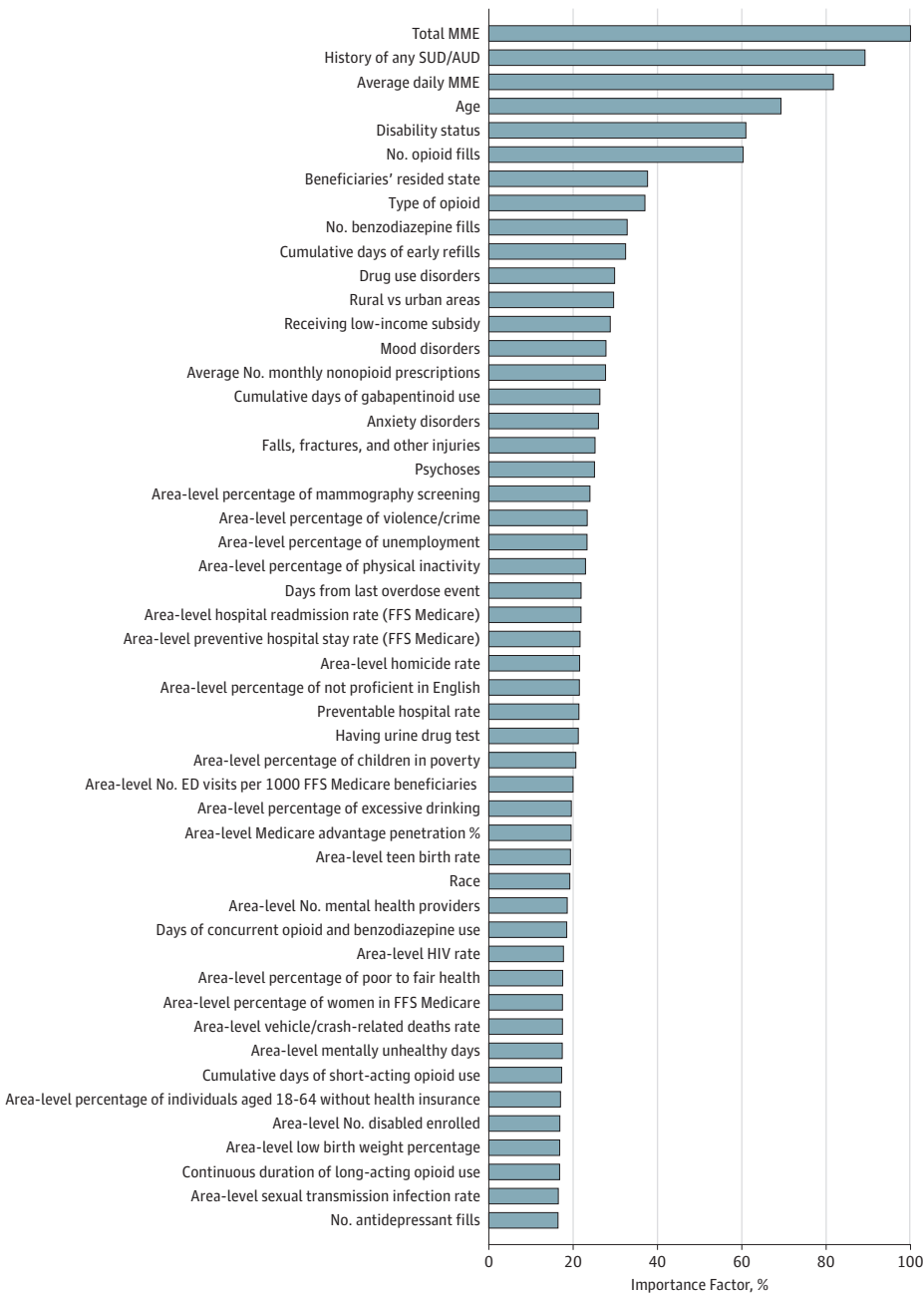


Table 2 compares the performance of the DNN algorithm measures with the Centers for Medicare & Medicaid Services opioid safety measures. By targeting the high-risk group, DNN's algorithm captured approximately 90% of 297 individuals with actual overdose episodes (NNE = 56) in a 12-month period, with 14 917 (8.9%) of the overall cohort being misclassified as having overdose, whereas the Centers for Medicare & Medicaid Services measures captured 30% of 297 individuals actual overdose episodes (NNE = 108), with 5.51% of the overall cohort being misclassified as overdose. The GBM and DNN algorithms performed similarly (eTable 7 in the Supplement). Sensitivity analyses using 6-month windows yielded similar C statistic with an improved PPV compared with using 3-month windows (eg, C statistic = 0.89; 95% CI, 0.87-0.90; PPV = 0.36% in DNN).

Figure 3. Top 50 Important Predictors for Opioid Overdose Selected by Gradient Boosting Machine



Rather than *P* values or coefficients, the gradient boosting machine reports the importance of predictors included in a model. Importance is a measure of each variable's cumulative contribution toward reducing square error, or heterogeneity within the subset, after the data set is sequentially split according to that variable. Thus, importance reflects a variable's significance in prediction. Absolute importance is then scaled to give relative importance, with a maximum importance of 100. For example, the top 10 important predictors identified from the gradient boosting machine model included total opioid dose (eg, >1500 morphine milligram equivalent [MME] during 3 months), diagnosis of alcohol use disorders or substance use disorders (AUD/SUD), mean daily opioid dose (eg, >32 MME), age, disability status, total number of opioid prescriptions (eg, >4), beneficiary's state residency (eg, Florida, Kentucky, or New Jersey), type of opioid use (eg, with mixed schedules), total number of benzodiazepine prescription fills (eg, >3), and cumulative days of early prescription refills (eg, >19 days). ED indicates emergency department; FFS: fee-for-service.

Discussion

Using national Medicare data, we developed machine-learning models with strong performance for predicting opioid overdose. The GBM and DNN models achieved high C statistic (>0.90) for predicting overdose risk in the subsequent 3 months after initiation of treatment with prescription opioids and outperformed traditional classification techniques. As expected in a population with very low prevalence of the outcome, the PPV of the models was low; however, these algorithms effectively segmented the population into 3 risk groups according to predicted risk score, with three-quarters of the sample in a low-risk group with a negligible overdose rate and more than 90% of individuals with overdose captured in the high- and medium-risk groups. The ability to identify such risk groups has important potential for policymakers and payers who currently target interventions based on less accurate measures to identify patients at high risk.

We identified 7 previously published studies of opioid prediction models, each focused on predicting a different aspect of opioid use disorder and not applying advanced machine learning. The studies predicted a 12-month risk of opioid use disorder diagnosis using private insurance claims^{13,22}; 2-year risk of clinical, electronic medical record–documented problematic opioid use in a primary care setting²⁴; 12-month risk of overdose or suicide-associated events using data from the Veterans Health Administration³¹; 6-month risk of serious prescription opioid–induced respiratory depression or overdose using data from the Veterans Health Administration and claims data from Insurance Management Services private insurance^{25,30}; and 2-year risk of fatal or nonfatal overdose using electronic medical record data.²⁸ These studies had several key limitations, including use of case-control designs unable to calibrate to population-level data with the true incidence rate of overdose; measuring predictors at baseline rather than over time; capturing only the first overdose episode; inability to identify complex or nonintuitive relationships (interactions) between the predictors and outcomes; and having suboptimal prediction performance (with a C statistic of up to 0.72 in

Table 2. Comparison of Prediction Performance Between Centers for Medicare & Medicaid Services Measures and Deep Neural Network Measures Over a 12-Month Period

Performance Metric	DNN Measures ^a			CMS Opioid Safety Measures ^b	
	Low Risk	Medium Risk	High Risk	Low- or No-Risk Opioid Use	High-Risk Opioid Use
Total, No. (%)	112 548 (67.5)	38 846 (23.3)	15 186 (9.1)	157 299 (94.4)	9281 (5.5)
Predicted score, median (range)	14.0 (2.1-46.5)	62.8 (46.5-81.9)	88.1 (81.9-99.7)	NA	NA
No. of actual overdose episodes (% of each subgroup)	7 (0.006)	21 (0.05)	269 (1.77)	210 (0.13)	87 (0.93)
No. of actual nonoverdose episodes (% of each subgroup)	112 541 (99.99)	38 825 (99.94)	14 917 (98.22)	157 089 (99.86)	9194 (99.06)
Sensitivity, %	0	100	100	0	100
PPV, %	NA	0.05	1.77	NA	0.93
NNE	NA	2000	56	NA	108
Specificity, %	100	0	0	100	0
NPV, %	99.99	NA	NA	99.86	NA
Overall No. of misclassified overdose episodes (% of overall cohort) ^c	7 (0.004)	38 825 (23.3)	14 917 (8.95)	210 (0.12)	9194 (5.51)
% of all overdose episodes captured over 12 mo (n = 297)	2.35	7.07	90.57	70.7	29.29

Abbreviations: CMS, Centers for Medicare & Medicaid Services; DNN, deep neural network; NA, not able to calculate; NNE, number needed to evaluate; NPV, negative predictive value; PPV, positive predictive value.

^a In contrast to Table 1, the measures were defined according to a 12-month period rather than a 3-month period. The sample size was smaller than in the main analysis because it required people to have at least 12 months of follow-up.

^b The 2019 CMS opioid safety measures are meant to identify high-risk individuals or utilization behavior.^{6,4} These measures include 3 metrics: (1) high-dose use, defined as higher than 120 morphine milligram equivalent (MME) for 90 or more continuous days, (2) 4 or more opioid prescribers and 4 or more pharmacies, and (3) concurrent opioid and benzodiazepine use for 30 or more days.

^c If classifying medium- and high-risk groups as overdose for DNN and low-risk group as nonoverdose, then individuals with actual nonoverdose in these 2 groups were misclassified. If classifying those with any of CMS high-risk opioid use measures as overdose, and the remaining group considered as nonoverdose, then individuals with actual nonoverdose in the high-risk groups were misclassified. The PPV and NNE were not able to calculate for the low-risk group because this group was considered as nonoverdose. Similarly, the NPV was not able to calculate for the medium- and high-risk groups because these groups were considered as overdose. Detailed definitions of prediction performance metrics are provided in eFigure 3 in the Supplement.

non-case-control designs). The present study overcomes these limitations using a population-based sample and machine-learning methods. To our knowledge, this study is the first to predict overdose risk in the subsequent 3-month period after initiation of treatment with prescription opioids as opposed to 1-year or longer period.

The extant literature in predicting health outcomes often focuses on C statistics rather than the full spectrum of prediction performance. This study found high C statistics (>0.90) from machine-learning approaches. However, although opioid overdose represents a particularly important outcome, it is a rare outcome, especially in the Medicare population. Relying on C statistics alone may lead to overestimating the advantages of a prediction tool or underestimating the costs of clinical resources involved. For a preimplementation evaluation of a clinical prediction tool, it is recommended that researchers report sensitivity and at least 1 other metric (eg, PPV, NNE, or estimated alert rate) to present a more complete picture of the performance characteristics of a specific model.^{59,61} In this study, the NNE value using DNN and GBM algorithms is similar to other commonly used cancer screening tests, such as annual mammography to prevent 1 breast cancer death (NNE = 233-1316, varying by subgroups with different underlying risk).⁶⁶

Unlike sensitivity and specificity, which are properties of the test alone, the PPV and NPV are affected by the prevalence of the outcome in the population tested. Low outcome prevalence leads to low PPV and high NPV, even in tests with high sensitivity and specificity, and could limit the clinical utility of a prediction algorithm such as ours because of false-positives. Other tests with good discrimination have low PPV because of overall prevalence, including trisomy 21 screening in 20- to 30-year-old women (prevalence of approximately 1:1200),⁶⁷ with a PPV of 1.7% at a test with sensitivity higher than 99% and specificity higher than 95%. Despite the low PPV in this study, our risk stratification strategies may more efficiently guide the targeting of opioid interventions among Medicare beneficiaries compared with existing measures. This strategy first excludes most (approximately 75%) prescription opioid users with negligible overdose risk from burdensome interventions like pharmacy lock-in programs and specialty referrals. Targeting medium- and/or high-risk groups can capture nearly all (90%) overdose episodes by focusing on only 25% of the population, which greatly frees up resources for payers and patients. For those in the high- and medium-risk groups, although most will be false-positives for overdose given the overall low prevalence, additional screening and assessment may be warranted. Although certainly not perfect, these machine-learning models allow interventions to be targeted to the small number of individuals who are at greater risk, and these models are more useful than other prediction criteria that have considerably more false-positives.

Limitations

The study has important limitations. First, patients may obtain opioids from nonmedical settings, which are not captured in claims data. Second, this study captured overdose episodes in medical settings and missed overdose episodes that occurred outside of medical settings, which are not captured in claims data. Third, the study relied on administrative billing data that lacked laboratory results and sociobehavioral information. This limitation can be addressed in the future with more robust linked data. In addition, although the study was novel in measuring overdose risk in the subsequent 3 months after initiation of prescription opioids, it used older data with complete claims capture; translation into real-time risk scores can be complicated by the lag in claims completion after the time of visit. Fourth, our focus was on predicting opioid overdose, and not opioid misuse, which is difficult to measure solely from claims data. Fifth, prediction algorithms and findings derived from the fee-for-service Medicare population may not generalize to individuals enrolled in Medicare Advantage plans or to other populations with different demographic profiles or programmatic features. However, the models may have better prediction performance in settings in which overdose is less rare (eg, Medicaid).

Conclusions

This study demonstrates the feasibility and potential of machine-learning prediction models with routine administrative claims data available to payers. These models have high C statistics and good prediction performance and appear to be valuable tools for more accurately and efficiently identifying individuals at high risk of opioid overdose.

ARTICLE INFORMATION

Accepted for Publication: February 1, 2019.

Published: March 22, 2019. doi:[10.1001/jamanetworkopen.2019.0968](https://doi.org/10.1001/jamanetworkopen.2019.0968)

Correction: This article was corrected on July 5, 2019, to fix data errors in Methods, Results, Table 1, Figure 2, and eAppendix and eTable 7 in the [Supplement](#).

Open Access: This is an open access article distributed under the terms of the [CC-BY License](#). © 2019 Lo-Ciganic W-H et al. *JAMA Network Open*.

Corresponding Authors: Wei-Hsuan Lo-Ciganic, PhD, Department of Pharmaceutical Outcomes & Policy, College of Pharmacy, University of Florida, PO Box 100496, 1225 Center Dr, HPNP Bldg, Room 3338, Gainesville, FL 32610-0496 (wlociganic@cop.ufl.edu).

Author Affiliations: Department of Pharmaceutical Outcomes & Policy, College of Pharmacy, University of Florida, Gainesville (Lo-Ciganic, Huang); Department of Mathematics, University of Arizona, Tucson (Zhang); Carnegie Mellon University, Heinz College, Pittsburgh, Pennsylvania (Weiss); Department of Health Outcomes & Biomedical Informatics, University of Florida, College of Medicine, Gainesville (Wu); Division of Rheumatology, Department of Medicine, and the University of Arizona Arthritis Center, University of Arizona, Tucson (Kwoh); Department of Health Policy and Management, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania (Donohue); Program for Addiction Research, Clinical Care, Knowledge, and Advocacy, Division of Epidemiology, Department of Internal Medicine, University of Utah, Salt Lake City (Cochran, Gordon); Informatics, Decision-Enhancement, and Analytic Sciences Center, Veterans Affairs Salt Lake City Health Care System, Salt Lake City, Utah (Gordon); Department of Pharmacy, Practice and Science, College of Pharmacy, University of Arizona, Tucson (Malone); Center for Pharmaceutical Policy and Prescribing, Health Policy Institute, University of Pittsburgh, Pittsburgh, Pennsylvania (Kuza, Gellad); Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania (Gellad); Center for Health Equity Research Promotion, Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, Pennsylvania (Gellad).

Author Contributions: Drs Lo-Ciganic and Huang had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Lo-Ciganic, Zhang, Kwoh, Donohue, Cochran, Gordon, Malone, Gellad.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Lo-Ciganic, Gellad.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Lo-Ciganic, Huang, Zhang, Weiss, Wu.

Obtained funding: Lo-Ciganic, Zhang, Kwoh, Donohue, Cochran, Gordon, Malone, Gellad.

Administrative, technical, or material support: All authors.

Supervision: Lo-Ciganic, Gellad.

Conflict of Interest Disclosures: Drs Lo-Ciganic, Huang, Zhang, Kwoh, Donohue, Cochran, Gordon, Malone, Kuza, and Gellad reported grants from the National Institutes of Health (NIH)/National Institute on Drug Abuse during the conduct of the study. Dr Lo-Ciganic reported grant support from the Pharmaceutical Research and Manufacturers of America Foundation during the conduct of the study. Dr Wu reported grants from the NIH during the conduct of the study. Dr Kwoh reported grant funding from AbbVie and EMD Serono; grants and personal fees from EMD Serono; and personal fees from Fidia, Regulus, GSK, Taiwan Liposome Company Inc, Kolon Tissue Gene, and Express Scripts outside of the submitted work. No other disclosures were reported.

Funding/Support: This study was supported by grant R01DA044985 from the NIH/National Institute on Drug Abuse and by a 2017 Pharmaceutical Research and Manufacturers of America Foundation Research Starter Award (Dr Lo-Ciganic).

Role of the Funder/Sponsor: The funders or sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: The views presented here are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs or the US government.

REFERENCES

1. Substance Abuse and Mental Health Services Administration Center for Behavioral Health Statistics and Quality. Results from the 2016 National Survey on Drug Use and Health: detailed tables. <https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.pdf>. Published September 15, 2017. Accessed February 15, 2019.
2. Centers for Disease Control and Prevention. National Center for Health Statistics, 2016: multiple cause of death data, 1999-2017. <http://wonder.cdc.gov/mcd.html>. Accessed January 27, 2019.
3. Rudd RA, Seth P, David F, Scholl L. Increases in drug and opioid-involved overdose deaths - United States, 2010-2015. *MMWR Morb Mortal Wkly Rep*. 2016;65(50-51):1445-1452. doi:10.15585/mmwr.mm655051e1
4. Seth P, Scholl L, Rudd RA, Bacon S. Overdose deaths involving opioids, cocaine, and psychostimulants - United States, 2015-2016. *MMWR Morb Mortal Wkly Rep*. 2018;67(12):349-358. doi:10.15585/mmwr.mm6712a1
5. Florence CS, Zhou C, Luo F, Xu L. The economic burden of prescription opioid overdose, abuse, and dependence in the United States, 2013. *Med Care*. 2016;54(10):901-906. doi:10.1097/MLR.0000000000000625
6. Roberts AW, Gellad WF, Skinner AC. Lock-in programs and the opioid epidemic: a call for evidence. *Am J Public Health*. 2016;106(11):1918-1919. doi:10.2105/AJPH.2016.303404
7. The US Congressional Research Service. The SUPPORT for Patients and Communities Act (P.L.115-271): Medicare Provisions. https://www.everycrsreport.com/files/20190102_R45449_231fb05ad093244bc8b91a84133fe310b2892e8e.pdf. Updated January 2, 2019. Accessed January 28, 2019.
8. Webster LR, Webster RM. Predicting aberrant behaviors in opioid-treated patients: preliminary validation of the Opioid Risk Tool. *Pain Med*. 2005;6(6):432-442. doi:10.1111/j.1526-4637.2005.00072.x
9. Ives TJ, Chelminski PR, Hammett-Stabler CA, et al. Predictors of opioid misuse in patients with chronic pain: a prospective cohort study. *BMC Health Serv Res*. 2006;6:46-55. doi:10.1186/1472-6963-6-46
10. Becker WC, Sullivan LE, Tetrault JM, Desai RA, Fiellin DA. Non-medical use, abuse and dependence on prescription opioids among U.S. adults: psychiatric, medical and substance use correlates. *Drug Alcohol Depend*. 2008;94(1-3):38-47. doi:10.1016/j.drugalcdep.2007.09.018
11. Hall AJ, Logan JE, Toblin RL, et al. Patterns of abuse among unintentional pharmaceutical overdose fatalities. *JAMA*. 2008;300(22):2613-2620. doi:10.1001/jama.2008.802
12. Centers for Disease Control and Prevention (CDC). Overdose deaths involving prescription opioids among Medicaid enrollees - Washington, 2004-2007. *MMWR Morb Mortal Wkly Rep*. 2009;58(42):1171-1175.
13. White AG, Birnbaum HG, Schiller M, Tang J, Katz NP. Analytic models to identify patients at risk for prescription opioid abuse. *Am J Manag Care*. 2009;15(12):897-906.
14. Dunn KM, Saunders KW, Rutter CM, et al. Opioid prescriptions for chronic pain and overdose: a cohort study. *Ann Intern Med*. 2010;152(2):85-92. doi:10.7326/0003-4819-152-2-201001190-00006
15. Edlund MJ, Martin BC, Devries A, Fan MY, Braden JB, Sullivan MD. Trends in use of opioids for chronic noncancer pain among individuals with mental health and substance use disorders: the TROUP study. *Clin J Pain*. 2010;26(1):1-8. doi:10.1097/AJP.0b013e3181b99f35
16. Sullivan MD, Edlund MJ, Fan MY, Devries A, Brennan Braden J, Martin BC. Risks for possible and probable opioid misuse among recipients of chronic opioid therapy in commercial and medicaid insurance plans: the TROUP Study. *Pain*. 2010;150(2):332-339. doi:10.1016/j.pain.2010.05.020
17. Bohnert AS, Valenstein M, Bair MJ, et al. Association between opioid prescribing patterns and opioid overdose-related deaths. *JAMA*. 2011;305(13):1315-1321. doi:10.1001/jama.2011.370
18. Volkow ND, McLellan TA, Cotto JH, Karithanom M, Weiss SR. Characteristics of opioid prescriptions in 2009. *JAMA*. 2011;305(13):1299-1301. doi:10.1001/jama.2011.401
19. Webster LR, Cochella S, Dasgupta N, et al. An analysis of the root causes for opioid-related overdose deaths in the United States. *Pain Med*. 2011;12(suppl 2):S26-S35. doi:10.1111/j.1526-4637.2011.01134.x
20. Cepeda MS, Fife D, Chow W, Mastrogianni G, Henderson SC. Assessing opioid shopping behaviour: a large cohort study from a medication dispensing database in the US. *Drug Saf*. 2012;35(4):325-334. doi:10.2165/11596600-000000000-00000

21. Peirce GL, Smith MJ, Abate MA, Halverson J. Doctor and pharmacy shopping for controlled substances. *Med Care*. 2012;50(6):494-500. doi:[10.1097/MLR.0b013e31824ebd81](#)
22. Rice JB, White AG, Birnbaum HG, Schiller M, Brown DA, Roland CL. A model to identify patients at risk for prescription opioid abuse, dependence, and misuse. *Pain Med*. 2012;13(9):1162-1173. doi:[10.1111/j.1526-4637.2012.01450.x](#)
23. Gwira Baumblatt JA, Wiedeman C, Dunn JR, Schaffner W, Paulozzi LJ, Jones TF. High-risk use by patients prescribed opioids for pain and its role in overdose deaths. *JAMA Intern Med*. 2014;174(5):796-801. doi:[10.1001/jamainternmed.2013.12711](#)
24. Hylan TR, Von Korff M, Saunders K, et al. Automated prediction of risk for problem opioid use in a primary care setting. *J Pain*. 2015;16(4):380-387. doi:[10.1016/j.jpain.2015.01.011](#)
25. Zedler B, Xie L, Wang L, et al. Development of a risk index for serious prescription opioid-induced respiratory depression or overdose in Veterans' Health Administration patients. *Pain Med*. 2015;16(8):1566-1579. doi:[10.1111/pme.12777](#)
26. Cochran G, Gordon AJ, Lo-Ciganic WH, et al. An examination of claims-based predictors of overdose from a large Medicaid program. *Med Care*. 2017;55(3):291-298. doi:[10.1097/MLR.0000000000000676](#)
27. Carey CM, Jena AB, Barnett ML. Patterns of potential opioid misuse and subsequent adverse outcomes in Medicare, 2008 to 2012. *Ann Intern Med*. 2018;168(12):837-845. doi:[10.7326/M17-3065](#)
28. Glanz JM, Narwaney KJ, Mueller SR, et al. Prediction model for two-year risk of opioid overdose among patients prescribed chronic opioid therapy. *J Gen Intern Med*. 2018;33(10):1646-1653. doi:[10.1007/s11606-017-4288-3](#)
29. Rose AJ, Bernson D, Chui KKH, et al. Potentially inappropriate opioid prescribing, overdose, and mortality in Massachusetts, 2011-2015. *J Gen Intern Med*. 2018;33(9):1512-1519. doi:[10.1007/s11606-018-4532-5](#)
30. Zedler BK, Saunders WB, Joyce AR, Vick CC, Murrelle EL. Validation of a screening risk index for serious prescription opioid-induced respiratory depression or overdose in a US commercial health plan claims database. *Pain Med*. 2018;19(1):68-78. doi:[10.1093/pm/pnx009](#)
31. Oliva EM, Bowe T, Tavakoli S, et al. Development and applications of the Veterans Health Administration's Stratification Tool for Opioid Risk Mitigation (STORM) to improve opioid safety and prevent overdose and suicide. *Psychol Serv*. 2017;14(1):34-49. doi:[10.1037/ser0000099](#)
32. Iams JD, Newman RB, Thom EA, et al; National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. Frequency of uterine contractions and the risk of spontaneous preterm delivery. *N Engl J Med*. 2002;346(4):250-255. doi:[10.1056/NEJMoa002868](#)
33. Hsieh E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS. Identifying important risk factors for survival in patient with systolic heart failure using random survival forests. *Circ Cardiovasc Qual Outcomes*. 2011;4(1):39-45. doi:[10.1161/CIRCOUTCOMES.110.939371](#)
34. Gorodeski EZ, Ishwaran H, Kogalur UB, et al. Use of hundreds of electrocardiographic biomarkers for prediction of mortality in postmenopausal women: the Women's Health Initiative. *Circ Cardiovasc Qual Outcomes*. 2011;4(5):521-532. doi:[10.1161/CIRCOUTCOMES.110.959023](#)
35. Chen G, Kim S, Taylor JM, et al. Development and validation of a quantitative real-time polymerase chain reaction classifier for lung cancer prognosis. *J Thorac Oncol*. 2011;6(9):1481-1487. doi:[10.1097/JTO.0b013e31822918bd](#)
36. Amalakuhan B, Kiljanek L, Parvathaneni A, Hester M, Cheriya P, Fischman D. A prediction model for COPD readmissions: catching up, catching our breath, and improving a national problem. *J Community Hosp Intern Med Perspect*. 2012;2(1):9915-9921. doi:[10.3402/jchimp.v2i1.9915](#)
37. Chirikov VV, Shaya FT, Onukwugha E, Mullins CD, dosReis S, Howell CD. Tree-based claims algorithm for measuring pretreatment quality of care in Medicare disabled hepatitis C patients. *Med Care*. 2017;55(12):e104-e112.
38. Thottakkara P, Ozrazgat-Baslanti T, Hupf BB, et al. Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications. *PLoS One*. 2016;11(5):e0155705. doi:[10.1371/journal.pone.0155705](#)
39. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63. doi:[10.7326/M14-0697](#)
40. Bossuyt PM, Reitsma JB, Bruns DE, et al; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527. doi:[10.1136/bmj.h5527](#)

41. Herzig SJ, Rothberg MB, Cheung M, Ngo LH, Marcantonio ER. Opioid utilization and opioid-related adverse events in nonsurgical patients in US hospitals. *J Hosp Med*. 2014;9(2):73-81. doi:10.1002/jhm.2102
42. Unick GJ, Rosenblum D, Mars S, Ciccarone D. Intertwined epidemics: national demographic trends in hospitalizations for heroin- and opioid-related overdoses, 1993-2009. *PLoS One*. 2013;8(2):e54496. doi:10.1371/journal.pone.0054496
43. Larochelle MR, Zhang F, Ross-Degnan D, Wharam JF. Rates of opioid dispensing and overdose after introduction of abuse-deterrent extended-release oxycodone and withdrawal of propoxyphene. *JAMA Intern Med*. 2015;175(6):978-987. doi:10.1001/jamainternmed.2015.0914
44. Fulton-Kehoe D, Sullivan MD, Turner JA, et al. Opioid poisonings in Washington State Medicaid: trends, dosing, and guidelines. *Med Care*. 2015;53(8):679-685. doi:10.1097/MLR.0000000000000384
45. Yang Z, Wilsey B, Bohm M, et al. Defining risk of prescription opioid overdose: pharmacy shopping and overlapping prescriptions among long-term opioid users in medicaid. *J Pain*. 2015;16(5):445-453. doi:10.1016/j.jpain.2015.01.475
46. Edlund MJ, Martin BC, Fan MY, Braden JB, Devries A, Sullivan MD. An analysis of heavy utilizers of opioids for chronic noncancer pain in the TROUP study. *J Pain Symptom Manage*. 2010;40(2):279-289. doi:10.1016/j.jpainsymman.2010.01.012
47. Lange A, Lasser KE, Xuan Z, et al. Variability in opioid prescription monitoring and evidence of aberrant medication taking behaviors in urban safety-net clinics. *Pain*. 2015;156(2):335-340. doi:10.1097/01.j.pain.0000460314.73358.ff
48. Jann M, Kennedy WK, Lopez G. Benzodiazepines: a major component in unintentional prescription drug overdoses with opioid analgesics. *J Pharm Pract*. 2014;27(1):5-16. doi:10.1177/0897190013515001
49. Liu Y, Logan JE, Paulozzi LJ, Zhang K, Jones CM. Potential misuse and inappropriate prescription practices involving opioid analgesics. *Am J Manag Care*. 2013;19(8):648-665.
50. Logan J, Liu Y, Paulozzi L, Zhang K, Jones C. Opioid prescribing in emergency departments: the prevalence of potentially inappropriate prescribing and misuse. *Med Care*. 2013;51(8):646-653. doi:10.1097/MLR.0b013e318293c2c0
51. Mack KA, Zhang K, Paulozzi L, Jones C. Prescription practices involving opioid analgesics among Americans with Medicaid, 2010. *J Health Care Poor Underserved*. 2015;26(1):182-198. doi:10.1353/hpu.2015.0009
52. Sharabiani MT, Aylin P, Bottle A. Systematic review of comorbidity indices for administrative data. *Med Care*. 2012;50(12):1109-1118. doi:10.1097/MLR.0b013e31825f64d0
53. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27. doi:10.1097/00005650-199801000-00004
54. Gordon AJ, Lo-Ciganic WH, Cochran G, et al. Patterns and Quality of Buprenorphine Opioid Agonist Treatment in a Large Medicaid Program. *J Addict Med*. 2015;9(6):470-477. doi:10.1097/ADM.0000000000000164
55. HipXChange. Area deprivation index datasets. <https://www.hipxchange.org/ADI>. Accessed November 13, 2018.
56. County Health Rankings and Roadmaps. Use the data. <http://www.countyhealthrankings.org/explore-health-rankings/use-data>. Accessed November 3, 2018.
57. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York, NY: Springer; 2008.
58. Chu A, Ahn H, Halwan B, et al. A decision support system to facilitate management of patients with acute gastrointestinal bleeding. *Artif Intell Med*. 2008;42(3):247-259. doi:10.1016/j.artmed.2007.10.003
59. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*. 2015;10(3):e0118432. doi:10.1371/journal.pone.0118432
60. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845. doi:10.2307/2531595
61. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care*. 2015;19:285-290. doi:10.1186/s13054-015-0999-1
62. Tufféry S. *Data Mining and Statistics for Decision Making*. West Sussex, UK: John Wiley & Sons; 2011. doi:10.1002/9780470979174
63. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J*. 2005;47(4):458-472. doi:10.1002/bimj.200410135

64. Centers for Medicare & Medicaid Services (CMS). Announcement of calendar year (CY) 2019 Medicare Advantage capitation rates and Medicare Advantage and Part D payment policies and final call letter. <https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/Announcement2019.pdf>. Accessed November 6, 2018.
65. Goldstein BA, Pencina MJ, Montez-Rath ME, Winkelmayer WC. Predicting mortality over different time horizons: which data elements are needed? *J Am Med Inform Assoc*. 2017;24(1):176-181. doi:10.1093/jamia/ocw057
66. Hendrick RE, Helvie MA. Mammography screening: a new estimate of number needed to screen to prevent one breast cancer death. *AJR Am J Roentgenol*. 2012;198(3):723-728. doi:10.2214/AJR.11.7146
67. Lutgendorf MA, Stoll KA. Why 99% may not be as good as you think it is: limitations of screening for rare diseases. *J Matern Fetal Neonatal Med*. 2016;29(7):1187-1189.

SUPPLEMENT.

eAppendix. Machine Learning Approaches Used in the Study

eTable 1. Diagnosis Codes for the Exclusion of Patients With Malignant Cancers Based on the National Committee for Quality Assurance (NCQA)'s Opioid Measures in 2018 Healthcare Effectiveness Data and Information Set (HEDIS)

eTable 2. Diagnosis Codes for Identifying Opioid Overdose

eTable 3. Other Diagnosis Codes Used to Identify the Likelihood of Opioid Overdose

eTable 4. Summary of Predictor Candidates (n=268) Measured in 3-month Windows for Predicting Subsequent Opioid Overdose

eTable 5. Opioid Overdose and Sociodemographic Characteristics Among Medicare Beneficiaries (n=560,057), Divided Into Training, Testing, and Validation Samples

eTable 6. Prediction Performance Measures for Predicting Opioid Overdose, Across Different Machine Learning Methods With Varying Sensitivity and Specificity

eTable 7. Comparison of Prediction Performance Using Any of Centers for Medicaid & Medicaid Services (CMS) High-Risk Opioid Use Measures vs. Deep Neural Network (DNN) and Gradient Boosting Machine (GBM) in the Validation Sample (n=166,580) Over a 12-month Period

eFigure 1. Sample Size Flow Chart of Study Cohort

eFigure 2. Illustrations of Two Study Designs: 3-month Windows for Measuring Predictor Candidates and Overdose Events

eFigure 3. Classification Matrix and Definition of Prediction Performance Metrics

eFigure 4. Prediction Performance Matrix Across Machine Learning Approaches in Predicting Opioid Overdose Risk in the Subsequent 3 Months: Sensitivity Analyses Including the Information Measured in All the Historical 3-Months Windows

eFigure 5. Scatter Plot Between Deep Neural Network (DNN) and Gradient Boosting Machine (GBM)'s Prediction Scores

eFigure 6. Top 50 Important Predictors for Opioid Overdose Selected by Random Forest (RF)

eReferences